图书情报工作
第 65 卷 第 24 期 2021 年 12 月

汉语语境下数字人文项目分析与启示
——2020 年数字人文年会（DH2020）项目评选综述

■ 沈立力　张宏玲　韩春磊　刘炜
上海图书馆上海科学技术情报研究所　上海 200031

摘　要：[目的/意义]对上海图书馆举办的 2020 年数字人文年会会议征集的 51 个汉语语境下数字人文项目进行分析，总结汉语世界关于中华文化主题的数字人文项目概况，为国内数字人文的未来发展提供参考。[方法/过程]采用网络调查法和数据分析法对 51 个数字人文项目进行分类研究，根据项目的研究对象、研究方法、基础设施类型和项目成果的现实意义进行分析与讨论以获取有价值的特征。就“数字人文与人文学科”“技术手段与方法应用”“基础设施建设”“服务模式与问题解决”四个维度进行深入探讨。[结果/结论]51 个项目的总结显示出五大特点：在历史学、文学领域蓬勃发展，机器学习与数字人文深度结合，数字资源建设至关重要，综合性数字人文平台的时代已经到来，跨学科、跨机构、跨国家的合作性项目开始显现。本文也针对当前数字人文项目存在的不足提出了促进机构间合作及多学科交叉融合，加强大数据交叉思维和人文素养培养，加强区域协同发展，继续提升学科服务能力和基础资源建设的主要建议。

关键词：数字人文　DH2020　数字人文年会　中华文化　项目分析
分类号：G250.7
DOI：10.13266/j.issn.0252－3116.2021.24.001

数字人文落地中国已整十年，十年间，国内数字人文领域一方面奋力追赶世界脚步，另一方面也在追问过去、梳理当下。以“积淀与超越：数字人文与中华文化（Benevolence and Excellence：Digital Humanities and Chinese Culture）”为主题的数字人文年会（DH2020）于 2020 年在上海召开。本届会议深度聚焦数字人文领域的中国问题，反映了中国数字人文领域的最新研究成果、学科产出和发展趋势，为国内外人文学者提供一个高水平、高层次的学术交流平台。会议期间发起了中国主题的数字人文优秀项目的征集和评选活动，此次大会征集到的项目集中代表了近十年来汉语世界数字人文的重要实践成果。本文依托各个项目详尽的申报材料，从研究对象、研究方法、基础设施类型分析和现实意义四个维度梳理归纳当前汉语语境下数字人文项目的进展现状，为我国数字人文研究与实践提供参考。

1　数字人文项目研究综述

数字人文（Digital Humanities，DH）起源于人文计算（Humanities Computing）[1]，随着计算机技术、互联网技术、多媒体技术等新兴技术的发展而形成新型跨学科研究领域[2]。20 世纪 90 年代，“人文计算”被广泛应用于超文本、音频、视频、数字地图等多个对象，涵盖历史、艺术、语言学等多个领域；2001 年，“数字人文”概念首次出现，并随着关联数据、本体等技术的出现得到了迅速发展[3]，呈现出数据密集型特征，其技术工具、数据资源和科研需求得到良性互动与协同发展[4]。

国际上，数字人文已有丰富成果。全球范围内有超过 183 个与“数字人文”相关的研究中心、数字人文协会、团队或实验室[4]，这些研究中心大多依托于大学或图档博机构，拥有完善的硬件设施和完备的科研能力。较为著名的研究中心包括耶鲁大学数字人文实验室[5]、麻省理工大学 HyperStudio[6]、日本立命馆大学京都文艺数字人文研究中心[7]、悉尼大学数字人文研究小组[8]；数字人文协会包括国际数字人文组织联盟（The Alliance of Digital Humanities Organizations）、澳大

作者简介：沈立力，馆员，硕士，E-mail：llshen@libnet.sh.cn；张宏玲，信息处理中心元数据制作部主任，副研究馆员，硕士；韩春磊，信息处理中心主任，副研究馆员；刘炜，副馆（所）长，研究员，博士，博士生导师。

收稿日期：2021－05－06　修回日期：2021－07－23　本文起止页码：4－17　本文责任编辑：杜杏叶

利亚数字人文协会(Australasian Association for Digital Humanities)、欧洲数字人文协会(European Association for Digital Humanities)、日本数字人文协会(Japanese Association for Digital Humanities)等。这些国外的研究中心和数字人文机构如火如荼开展了大量数字人文项目,相关实践经验值得参考和借鉴,因此我国学者对国际数字人文项目做了一些梳理和总结。邓要然[9]对14 个美国图书馆领域数字人文项目进行研究,总结其主题热点和特征。赵雪芹等[10]从项目设立、技术运用、展览服务等角度探索包括时间线旅行线项目(麻省理工学院)、纽伦堡实验项目(哈佛大学)、蓝山项目(普林斯顿大学)在内的美国排名前 10 位的多个高校图书馆数字人文项目。徐彤阳、顾婷婷[7]梳理日本数字人文项目整体概况,总结其在研究主题、研究深度与广度、数据开放程度以及研究重点方面的特征。徐彤阳、杨明睿[11]分析澳大利亚数字人文项目后发现澳大利亚数字人文项目具有以政府机构为指导,以高校图书馆为项目实施主力,以企业为保障的体系特点。林泽斐[12]以 DHcommons 项目数据库为实证,分析英国数字人文项目的研究热点。由此可见,我国学术界实时跟踪国外数字人文项目发展,特别是欧美、日本等代表性国家的数字人文科研动态,同时也为我国的数字人文项目工作提供借鉴。

国际数字人文项目建设的蓬勃发展也推动了国内数字人文研究的进展。自 2011 年国内第一个数字人文研究中心在武汉大学成立以来,国内高校和图档博机构纷纷成立相关的数字人文研究及工作机构,如北京大学数字人文研究中心、上海图书馆历史人文大数据中心、复旦大学人文社科数据研究所、华东师范大学调查与数据中心、南京大学高研院数字人文创研中心等,这些机构承担起了国内数字人文项目建设与发展的重任。国内学者也对这些项目的发展进行追踪和分析。刘炜、叶鹰[2]从宏观层面对数字人文整体方法论和理论结构进行研究探讨。高胜寒等[4]通过文献计量方法对国内数字人文已有成果进行综述,详细分析了该领域内知识基础、研究热点和研究趋势。蔡迎春[13]对国内特色资源数据库建设中的数字人文应用进行了研究。苏敏[14]侧重对国内图书馆开展的数字人文服务的研究成果进行论述,明确我国图书馆数字人文服务研究的发展方向。吴丽萍[15]运用文献计量法对图书馆在数字人文服务中的角色定位、研究工具、教育培训、岗位培训等内容进行评述。朱本军、聂华[3,16]在对第一第二届北京大学数字人文论坛会议进行综述时涉及了对部分国内数字人文项目的分析与梳理。吕璐成、韩涛[17]同样在 2019 年图书馆前沿技术论坛的会议综述中介绍了部分的数字人文项目。由此可见,目前国内数字人文的综述研究主要集中在对方法论和理论结构的探讨、发展方向和前景的讨论、图书馆数字人文服务的开展以及资源数据库和平台的建设,少有对汉语语境下数字人文项目全面和专题的综述,而已有的项目综述仅包含在历届会议综述中。因此,本文将依托 2020 年数字人文年会(DH2020)项目征集评选的契机,依据每个项目组提交的项目资料(包括系统平台工具建设、资源库、知识库建设及可视化等相关的网站、App、小程序、图片及视频等),对汉语语境下数字人文项目进行全面梳理和分析,从项目研究对象、研究方法、类型分析和现实意义四个方面对国内数字人文实践情况进行归纳总结评述。

2 DH2020 征集项目基本情况

2.1 项目评选情况

2020 数字人文年会优秀项目评选采取专家征集和项目组自行报名相结合的方式,共产生 76 个候选项目,组委会对 76 个候选项目中评选意向不确定的项目进行相关联系信息的初步确定后,发送了优秀项目评选意愿征询函,经过汇总最终产生 26 个项目参加优秀项目的评选。此外,对未参评的项目进行再次梳理,整理出 25 个数字人文项目,共计 51 个项目加入此次年会综述。优秀项目评选设置最佳项目奖、最佳创意奖、最佳美工设计奖和最佳题材奖 4 类奖项,最终评出 11 个优秀项目。项目基本情况见表 1 和表 2。

2.2 项目基本情况分析

从项目的建设机构所在国家、地区来分析,大多数汉语语境的数字人文项目建设方为国内机构,有 3 个项目有海外机构(美国)参与,另有 1 个来自德国的独立项目。而从图 1 所有国内独立建设的项目中分析,约 63%(33 个)的项目来自于北京(16 个)和上海(17 个),这反映出北京和上海是目前国内数字人文研究较为领先的地区,产出了较多优秀的研究成果,研究成果得到专家学者的肯定和认同。同时发现,台湾地区也十分重视数字人文项目的建设,共有 6 个项目被推荐参评,其项目建设成果受到领域学者的普遍认可。

从图 2 项目建设机构的类型来分析,51 个项目共有 33 家机构参与建设,其中 19 家为高校(占比约 58%),5 家科研机构(占比约 15%),5 家公共图档博(占比约 15%)和 4 家商业机构(占比约 12%),此外还

表 1 26 个参评项目基本情况

编号	项目名称	地区	建设机构	学科	项目类型	资源类型	内容覆盖
XM01	《中国历史人物资料库》(CBDB) 在线查询系统第二版——面向用户需求的重新设计与实现	北京	北京大学	史学	平台与工具、系统建设	结构化数据	广
XM02	唐宋文学编年地图	武汉	中南民族大学	文学、史学	数字资源建设	结构化数据	一般
XM03	董其昌数字人文展示系统	上海	上海博物馆	艺术	数字资源建设	图像、文本	窄
XM04	中国多世代人口数据库	香港	香港科技大学	史学、社会学	数字资源建设、系统建设	结构化数据	广
XM05	台湾大学数字人文中心建构之数字人文分析系统与个人 DH 研究平台	台湾	台湾大学	图书情报	平台与工具、系统建设	文本、结构化数据	广
XM06	六朝建康城市历史信息系统	南京	南京大学	史学	数字资源建设	结构化数据	窄
XM07	印藏:复旦大学印谱文献虚拟图书馆	上海	复旦大学	艺术	数字资源建设	图像	窄
XM08	南京地区侵华日军慰安所的 AR 故事地图	南京	南京大学	史学	数字资源建设	综合	窄
XM09	方志物产知识库构建及深度利用研究	南京	南京农业大学	史学	数字资源建设	文本、结构化数据	窄
XM10	基于 IIIF 的敦煌壁画数字叙事系统	武汉	武汉大学	艺术	数字资源建设	图像、文本	窄
XM11	高迁古村数字记忆网站	北京	中国人民大学	史学	数字资源建设、系统建设	综合	窄
XM12	“战疫中的复旦”专题数字特藏平台	上海	复旦大学	图书情报、医学	数字资源建设	综合	窄
XM13	历史文化名人游学足迹知识图谱的构建与可视化	上海	上海大学	史学	平台与工具	结构化数据	窄
XM14	中国古籍基础数据分析平台	上海	上海外国语大学	文学	数字资源建设、平台与工具	结构化数据、文本	广
XM15	多源异构学术成果大数据的融合与揭示	上海	复旦大学	图书情报	平台与工具	结构化数据	广
XM16	融合机器智能与群体智慧的古籍数据化实践	上海	华东师范大学	图书情报、文学	平台与工具	结构化数据、文本	广
XM17	中国近现代书画印本数据库	上海	华东师范大学	艺术	数字资源建设	图片	一般
XM18	分布式数字人文研究与教学实训环境	上海	华东师范大学	图书情报	平台与工具	结构化数据	一般
XM19	复旦大学东亚语言数据中心	上海	复旦大学	语言学	数字资源建设	结构化数据	窄
XM20	敦煌壁画主题词表及关联数据发布服务平台	武汉	武汉大学	艺术、史学	数字资源建设	结构化数据	窄
XM21	宋元学案知识图谱可视化系统	北京	北京大学	文学	数字资源建设	图谱	窄
XM22	丝绸之路地理信息开放平台	北京 陕西	首都师范大学 陕西师范大学出版社有限责任公司	史学	数字资源建设	结构化数据	窄
XM23	家谱知识服务平台	上海	上海图书馆上海科学技术情报研究所	史学	平台与工具、系统建设	文本、图像	广
XM24	SinoPedia:关联数据服务平台	上海	上海图书馆上海科学技术情报研究所	图书情报	方法和工具	结构化数据	窄
XM25	中国古代皇室家族树(赵宋、李唐、朱明)	北京	清华大学	史学	数字资源建设	文本	窄
XM26	中国历史地理信息系统(CHGIS)	上海 美国波士顿	复旦大学 哈佛大学	史学	数字资源建设、系统建设	结构化数据	广

表2　25个未参评项目基本情况

编号	项目名称	地区	建设机构	学科	项目类型	资源类型	内容覆盖
XM27	中国记忆	北京	国家图书馆	史学	数字资源建设	综合	广
XM28	北京记忆	北京	中国人民大学	史学	数字资源建设	综合	一般
XM29	国学宝典	北京	首都师范大学 北京国学时代文化传播股份有限公司	文学	数字资源建设	结构化数据、文本	广
XM30	丝绸之路地理信息系统项目	上海	复旦大学	史学	数字资源建设	结构化数据	窄
XM31	链上的开放知识图谱联盟	北京	中国中文信息学会语言与知识计算专业委员会	图书情报	平台与工具	图谱	广
XM32	学术地图发布平台	浙江 美国波士顿	浙江大学 哈佛大学	史学	数字资源建设、平台与工具	结构化数据	一般
XM33	数字敦煌	敦煌	敦煌研究院	史学	数字资源建设	结构化数据、图片	一般
XM34	清季职官表查询系统	台湾	台湾大学	史学	数字资源建设	结构化数据	窄
XM35	台湾百年历史地图	台湾	台湾“中央研究院”	史学	数字资源建设	结构化数据	窄
XM36	文化资源地理信息系统	台湾	台湾“中央研究院”	文学、史学	数字资源建设	结构化数据	窄
XM37	中华文明之时空基础架构（CCTS）	台湾	台湾“中央研究院”	史学	数字资源建设	结构化数据	窄
XM38	台湾历史文化地图（THCTS）	台湾	台湾“中央研究院”	史学	数字资源建设	结构化数据	窄
XM39	中国历代人物印鉴数据库	浙江	浙江图书馆	艺术	数字资源建设	图片	一般
XM40	宋词缱绻，何处画人间	浙江 北京	浙江大学 新华网数据新闻部	文学	数字资源建设	结构化数据、文本	窄
XM41	九歌——计算机古诗作诗系统	北京	清华大学	文学	平台与工具	结构化数据	窄
XM42	唐诗别苑	北京	北京师范大学	文学	数字资源建设	结构化数据	一般
XM43	中国地方历史文献数据库（契约文书）	上海	交通大学	史学	数字资源建设	文本	广
XM44	云南省古籍数字图书馆	云南	云南省图书馆	文学	数字资源建设	文本	广
XM45	再现圆明园	北京	北京清华城市规划设计研究院	史学	数字资源建设	综合	窄
XM46	历史人文大数据平台	上海	上海图书馆上海科学技术情报研究所	史学、文学、图书情报	数字资源建设、系统建设、平台与工具	文本、结构化工具	广
XM47	地方志数据库计划	德国	马克斯·普朗克科学史研究所	史学	平台与工具	结构化数据	窄
XM48	民国时期期刊语料库	芝加哥 上海	芝加哥大学文本光学实验室 上海图书馆上海科学技术情报研究所	史学	数字资源建设	结构化数据、文本	窄
XM49	中国历代人物传记资料库（CBDB）	美国波士顿 北京	哈佛大学费正清中国研究中心、“中央研究院历史语言研究所” 北京大学中国古代史研究中心	史学	数字资源建设	结构化数据	广
XM50	中国知网数字人文研究平台	北京	同方知网北京技术有限公司	图书情报	数字资源建设、平台与工具	综合	广
XM51	古籍整理平台	北京	中华书局	文学、史学	数字资源建设	结构化数据	广

有新华网等媒体机构参与其中。由此可见，高校依旧是数字人文项目建设和推行的“主力军”，科研机构和公共图档博机构则是该领域建设的有生力量，而商业机构让数字人文领域主体发展呈现跨行业、跨机构的新态势，助力数字人文研究成果的落地推广。另外需要指出的是，越来越多的高校，如北京大学、武汉大学、中国人民大学等，成立了数字人文研究中心，并承担起了项目建设者的角色，对数字人文项目进行研究和孵化，这些中心的建立表现出高校对数字人文这门交叉学科的重视，随着各高校数字人文研究中心的逐步建立和完善，将来必将涌现出更多优秀研究项目。

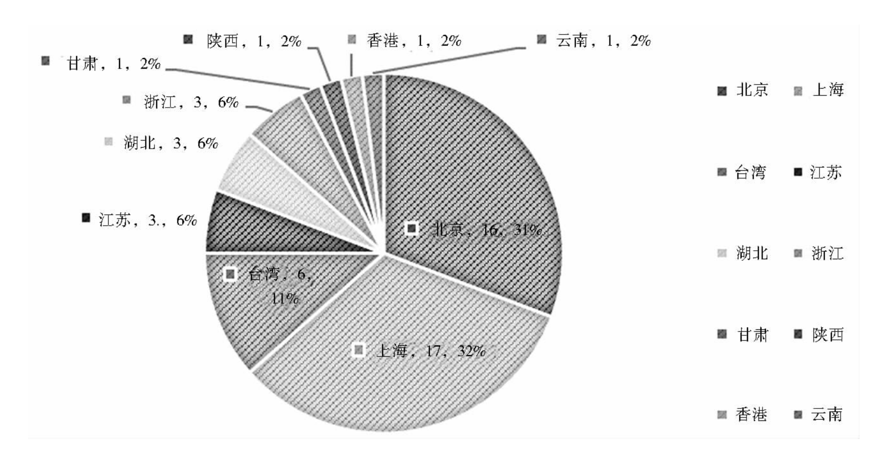


图1 国内数字人文项目建设地区分布(含合作建设机构)

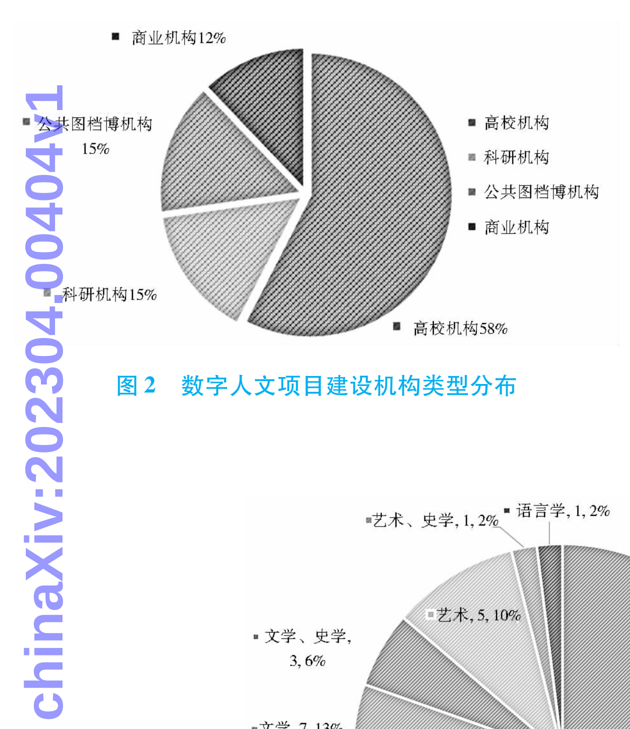


图2 数字人文项目建设机构类型分布

从图3项目涉及的学科分布中可以看到，51个项目中有24个项目的研究对象是历史学(占比47%)、7个项目是文学(占比13%)、5个项目是艺术(占比10%)。由此可见，历史学和文学仍然是数字人文领域研究的热门学科，而艺术领域则是数字人文新的关注对象。此外，从学科分布中还可以发现，有9个项目涉及了多个学科领域，数字人文研究在人文研究与数字技术的跨学科结合中起到了融合作用。最后，值得注意的是，跨学科的项目中有4个项目都是由图书情报学科参与或主导，可见图书馆和图情系是数字人文项目的中坚力量。

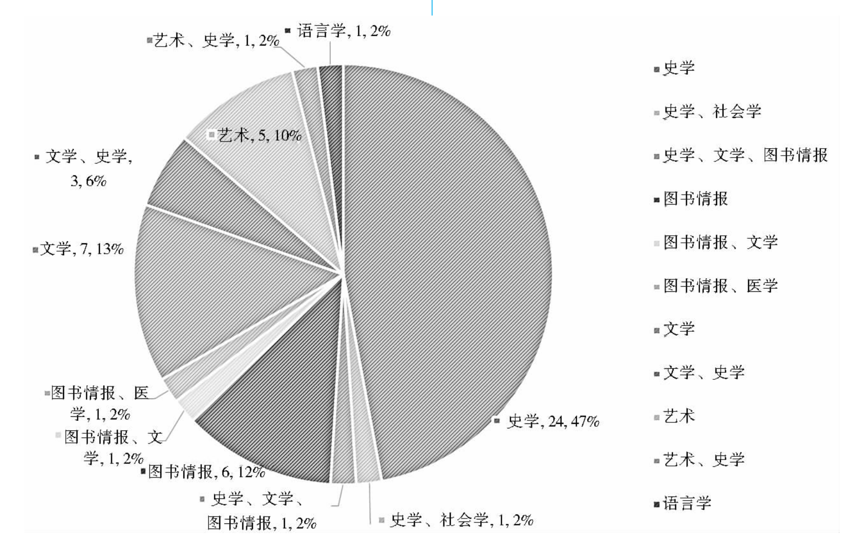


图3 数字人文项目学科分布

3 数字人文与人文学科：数字人文项目的研究对象

数字人文项目是数字技术在人文领域的实践与融合。人文学科研究正在向以数据为驱动的新型模式转型，形成"数字人文"研究新范式[18]。各人文学科领域的研究者和研究机构，包括美术馆、档案馆、博物馆(GLAM)等文化记忆机构都在建设数字人文基础设施，以支撑新的研究范式、知识生产、交流模式[19]。数字人文项目分布的学科主要为历史学、文学、艺术、图书情报、语言学及社会学等，本文根据各个项目的研究内容和研究对象，结合数字技术方法，对数字史学、数字文

学、数字艺术等细分领域的研究成果进行梳理归纳。

3.1 历史学领域

51个项目中历史学领域项目占比最大，通过对项目主题和内容的梳理，可分为文化遗产数字化进程、史学文献数据规模化和语义化建设、历史时空史料的研究和应用3个方向。这些项目不仅为传统的史学研究带来了新技术和新方法，也深化了史学研究内核、拓宽了外延。

3.1.1 文化遗产的数字化进程

2020年5月中宣部在《关于做好国家文化大数据体系建设工作通知》中明确指出要“分类采集梳理文化遗产数据，对全国公共文化机构、高等科研机构和文化生产机构各类藏品数据，标注中华民族文化基因，把非物质文化遗产记录成果中蕴含的优秀传统文化的精神标识提炼出来”[20]，强调了数字遗产项目建设的重要性和必要性。国内文化遗产数字化的项目很多，文化遗产的数字化工作离不开各类文化机构及社会各界的共同努力。此次项目征集中共有5个项目涉及文化遗产数字化主题，从建设机构、文化遗产类型、主要数字资源、主要技术、应用场景5个角度总结了相关特征。如表3所示：

表3 文化遗产数字化项目

项目名称	建设机构	文化遗产类型	主要数字资源	主要技术	应用场景
高迁古村数字记忆网站	中国人民大学数字人文研究中心	非物质文化遗产	建筑空间视觉资源、数字家谱、口述史、照片、捷报、门匾楹联、绘画作品、高迁村史、家训诵读录音	领域本体、命名实体识别	前站后库、以“记忆高迁 爱得我所”门户网站作为高迁乡土文化与历史知识的呈现窗口
数字敦煌	敦煌研究院	物质文化遗产	洞窟信息、壁画信息、音视频节目	VR技术、高清扫描	构建多元化与智能化相结合的石窟文物数字化资源库、建立数字资产管理系统和数字资源科学保障体系
再现圆明园	北京清华城市规划设计研究院	物质文化遗产	考古遗址现场记录、四十景图等数字化档案、历次测绘图	Re-relic系统、VR技术、3D全景建模	对建筑遗产进行虚拟再现和综合信息管理、支撑学术研究、考古发掘、保护工程、日常管理、民众参与
中国记忆	国家图书馆	非物质文化遗产	口述史料、影像史料、手稿、信件、照片、实物	专题库建设	整理中国现当代重大事件、重要人物专题文献，形成多载体、多种类的专题文献资源集合
北京记忆	中国人民大学人文北京研究中心	非物质文化遗产	图片、地方文献、资源地图、音视频	专题库建设	构建和呈现“老北京”的历史魅力与“新北京”的精神风貌，构建数字化北京记忆专题库

从表3中可见，文化遗产数字化项目的建设主体既包括公共及高校图书馆，也有敦煌研究院这样的科研机构。文化遗产数字化项目具体内容依据文化遗产类型可以分为物质文化遗产（包括保护型历史文物、历史建筑和人类文化遗址等）和非物质文化遗产（包括口头传统和表现形式、表演艺术和传统手工艺等）[21]。图书馆为主导的文化遗产数字化大多侧重于非物质文化遗产的数字化工作，具体而言就是在馆藏资源建设的基础上对特定主题资源利用命名实体识别、本体建设、专题数据库建设等技术对口述史料、家谱、音视频资料、手稿、信件等非物质文化遗产进行建设、组织、管理、保存、利用，从而达到对文化习俗、地域文化、口头传承、精神信仰等非物质文化遗产的记录、保护与传承的目的。而科研机构为主导的项目则大多是对物质文化遗产进行数字化加工，通过VR/AR技术、高清扫描、3D全景建模、地图绘制等技术手段对文化遗址、历史文物、历史建筑（群）等有形的文化遗产进行测量、标记、拍摄，并且进一步通过在线可视化为用户提供云端游览等新的文旅体验，从而达到虚拟还原、真实展现及保护传承的目的。

相关学者指出文化遗产数字化项目存在文化性、历史传承性、分散性、繁杂性和传播性5个专指特征[22]。而冯惠玲也在高迁古村数字记忆项目中指出，整体性、跨时空、活态化、广视域、多维叙事、语义化、传统文本的数字诠释以及沉浸式是当今文化遗产数字化项目前进的新方向。文化遗产数字化使得不可移动文物突破物理空间和地域局限，走出文物本体环境而呈现于公众，以虚拟和现实相结合、艺术与科技相融合方式，打造全新数字文化记忆[23]。

3.1.2 史学文献数据规模化与语义化建设

对史学文献进行大规模数字化加工，并在此基础上进行语义化处理是数字史学建设的基础性工作。其主要工作是将分散在史料中的人名、地名、物产名、著作、事件名等进行标注并转化为结构化数据，从而形成大规模知识库和数据集，为后续人文学者的地理空间分析、社会网络分析和文献计量分析等工作提供数据

基础。例如香港科技大学李中清、康文林团队的"中国多代人口数据库(CMGPD)",通过挖掘和建设基于户籍和族谱等中国历史档案资料的大规模历史数据库,实现了对长期人口记录和跨越多世代家庭变化的追踪。台湾大学数字人文中心的"清季职官表查询系统",通过对《清季职官表附人物录》的数字化,构建了涵盖乾隆 51 年至宣统 3 年的官职、人名、组织官制、上下属关系等信息的数据系统。上海交通大学图书馆建设的《中国地方历史文献数据库》收录约 35 万件地方历史文献,构建了规模最大的可全文检索的写本文献数据库。上海图书馆建设的"家谱知识服务平台"通过关联数据技术和知识图谱技术提供了丰富的语义化家谱数据,同时结合人名规范库、地名表、历史纪年表等关联开放数据,建成了全量级、细粒度、长时间、大空间、数据开放的全球性家谱数据库。

3.1.3 历史时空史料的研究和应用

历史地理项目建设是史学领域数字人文项目的另一个研究热点。此类项目往往通过结合历史史料和地理信息系统技术(Geographic Information System, GIS)辅助历史地理数字化地图的构建,多角度揭示并模拟特定时空序列下特定历史事件的演变过程,以及为国家相关课题的研究提供科研地理数据。南京大学的"南京地区侵华日军慰安所的 AR 故事地图"利用 GIS 技术和 AR 增强现实技术,将侵华日军设立的慰安所按时空顺序呈现在 AR 三维地图上,并将其与慰安所的历史资料相结合,制成南京地区侵华日军慰安所的 AR 故事地图。在丝绸之路这一史学主题的研究上,首都师范大学历史学院开发的"丝绸之路历史地理信息开放平台"将分散的历史文献资料提取成带有空间属性的历史数据,并与考古资料相结合构建了涵盖城镇、交通、文化、土地垦殖、生态环境等人文、科研数据的可视化地图数据库,为国家提供了丝绸之路的科研信息咨询数据。复旦大学侯杨方团队的"丝绸之路地理信息系统"则通过 WebGIS 复原了法显、玄奘取经、高仙芝远征等多条线路,并对沿线的古城分布情况进行了标注,首次实现了对传统丝绸之路在现代公路的地理位置上的精确复原。此外,台湾地区"中央研究院"的"台湾历史文化地图(THCTS)""文化资源地理信息系统""中华文明之时空架构(CCTS)"系列项目都是利用 WebGIS 技术将 400 年来台湾历史、文化、自然资源的地图信息建立在一个时空应用框架内,展示台湾地区的历史发展和环境变迁。

3.2 文学领域

文学领域项目按照对数据资源数字化加工程度可以划分为文本处理与基础语料建设、文本分析与知识探索两个方面。

3.2.1 文本处理与基础语料建设

在文学领域,数字技术应用最直接和最基础的工作就是文本的数字化和基础语料数据库的建设。从项目调研中发现,古籍文献资料数字化项目在对实体文献的扫描项目中占比最大。例如,首都师范大学研发了大型古籍文献全文检索数据库《国学宝典》,中华书局古联公司推出的古籍整理平台聚合了一系列数字古籍产品,如《中华善本古籍数据库》、《中华文史学术论著库》等。而云南省图书馆建设的云南省古籍数字图书馆则完成从宋朝至民国时期刻本、抄本、写本等不同版本的古籍数字化工作。

3.2.2 文本分析与知识探索

在对文献进行文本数字化加工的基础上,许多项目利用知识图谱、文本分析、全文自动识别等技术手段,将文献数据转换为机器可读的数据,并对文本内容进行深层次挖掘、对其中人、地、时、事及关系进行揭示。

在诗词领域,新华网数据新闻联合浙江大学可视化小组研究团队的"宋词缱绻,何处画人间"项目以《全宋词》为样本,通过知识图谱、社会网络分析等手段构建完成了宋代词人游历路线图、宋代词人生平及所处年代图谱。同时,以文本分析进行《全宋词》词频统计和常见意象统计,为宋词赏析提供新的数据化角度。北京师范大学的"唐诗别苑"平台则聚焦唐代诗人及其作品,使用自然语言处理技术实现对诗歌主题、情感、诗人的大规模知识挖掘,打造基于知识图谱的全唐诗语义检索与可视化平台。另外,清华大学的"九歌——计算机古诗作诗系统"则是采用最新的深度学习技术,结合多个专门设计的诗歌生成模型,基于超过 80 万首人类诗人创作的诗歌进行训练学习,研发了人工智能诗歌写作系统,引起人们对诗词创作的兴趣。

在近代报刊领域,芝加哥大学文本光学实验室和上海图书馆合作的"民国时期期刊语料库"通过语义检索、分类算法和文本重用等技术,对民国时期部分期刊进行元数据搜集、主题标引及文本 OCR 识别工作,并开发出相应的文本分析研究工具,为人文学者阅读和分析大规模文献提供多种尺度的"透镜",以达到在远读和细读的交替中发现各种抽象的文化模式的目的。而上海图书馆开发的"历史人文大数据平台"则

集合近代报纸、期刊、图书、家谱、手稿等特色馆藏资源，利用关联数据、知识图谱、IIIF、机器学习、GIS等新技术支撑计量统计、文本分析、社会网络关系分析、时空分析、3D建模等众多数字人文研究方法和范式，支撑不同人文学科研究的知识交流模式和应用场景。

在古籍领域，上海外国语大学建设的“中国古籍基础数据分析平台”在汇集古籍文字、文本、词典、典籍知识等中国古籍基础数据基础上为人文学者研究提供文本分析工具和中国存世典籍的目录检索和知识图谱服务。华东师范大学调查与数据中心建设的“融合机器智能与群体智慧的古籍数据化实践”项目则是通过人工智能技术在数字人文中的应用和众包协作模式，提供古籍全文自动识别、自动句读、图文自动标引等古籍数据化工作中需要用到的方法和工具。中南民族大学开发的“唐宋文学编年地图”项目则是解决了中国古代文学研究中长期存在的“资料离散”和“时空分离”两大难题，并首次提出“系地”概念，实现了文学信息的时空一体化呈现，客观还原了文学发生的场景，实现了GIS与文学的互动。

3.3 艺术领域与视觉传达

在对51个项目分析中发现数字人文在艺术领域逐渐发挥作用。印藏、壁画、书画的数字化呈现为人文艺术研究提供了极大便利。同时，越来越多的数字人文项目开始更多关注数据可视化呈现和视觉传达。

3.3.1 数据可视化

数据可视化是现代科技与人文艺术融合发展的产物，随着数字技术的发展，图像和数据的结合形式产生了变化，并得到了发展。数据可视化也为数字人文领域的研究提供了新视角。清华大学向帆团队设计建设的中国古代皇室家族树(赵宋、李唐、朱明)以中国古代人物关系数据库(CDBD)中的家族关系数据为基础，运用算法和可视化手段将个人家族关系进行相互判断、连接，并按时间顺序构成生长的家族树，从中发现血脉关系中的疑点。中国古代皇室家族树的构建工作既揭示了历史中大型家族的内部特征，又提出了不同于传统视觉的新问题。

3.3.2 视觉传达与用户体验

一些数字人文项目除了在数据可视化领域有新的尝试，在视觉传达和用户体验方面也不断吸收先进UI设计理念，以起到更明确揭示项目主题，更舒适交互体验，更清晰演绎数据模型、更直观视觉化转化的作用。中国人民大学的“高迁古村数字记忆网站”首页采用国内比较少见的全屏网页结构，每屏只向观众展示一个栏目的引言和入口，表现出更强的陈列感，能够有效地避免其他栏目的信息干扰，为访客提供沉浸式的浏览体验，让观众能够迅速融入到古色古韵的古村落文化氛围中。上海图书馆设计开发的“历史人文大数据平台”也同样采用了全屏网页结构，并在首页插入视频动画，使首页的展示更加直观和富有律动。同时，采用多屏下拉的方式向用户展示海派品牌、电影记忆、红色旅游等多个故事性场景，循环且动态地向用户展示历史人文大数据平台丰富的资源内容和特色，交互体验贯穿整个故事性场景的浏览、输入、输出过程。

3.3.3 印藏、壁画、书画的数字化呈现

在本次项目分析中发现，数字人文技术在书画、印藏等传统艺术领域有所作为，为这些领域的知识组织方式和展现形式带来了新方向。上海博物馆开发建设的董其昌数字人文展示系统首次将机器学习引入书画研究，依靠数据关联和量化分析，以可视化的形式展现了与董其昌相关的时、地、人、事，实现书画 精看，还原了一个以董其昌为中心而构筑的晚明时代的艺术生态圈，是对博物馆数据可视化的新尝试。复旦大学建设的“印藏：复旦大学印谱文献虚拟图书馆”引入国际图像互操作框架(International Image Interoperability Framework，IIIF)，采用Serverless架构，构建高性价比的大规模图片资源服务，重新组织并展示了金石学者松荫轩主人林章松的印藏藏品。华东师范大学建设的“中国近现代书画印本数据库”则是以图像资源为主要建设对象，运用IIIF、关联数据等技术手段，在展示董其昌、冯超然等书画名家的作品的同时，提供了内容注释、智能推荐、时空分析等书画作品欣赏、研究的新功能。武汉大学开发及建设了两个基于敦煌壁画主题的项目。一是“基于IIIF的敦煌壁画数字叙事系统”，二是“敦煌壁画主题词表及关联数据发布服务平台”。前者是以敦煌莫高窟第61窟中知名的《五台山图》壁画为案例，利用IIIF技术手段，引入数字叙事的理念，从人、事、时、地、物五个维度对壁画中的概念与实体进行解构和场景揭示。而后者则是构建了一个规范、全面的敦煌壁画领域主体词表，为敦煌壁画数字资源的深度语义标注、知识组织、信息关联与共享等提供一套受控词表，促进了敦煌壁画数字人文应用的开发。

3.4 图书情报领域与相关交叉学科

本次数字人文的很多项目都是图书情报学科领域内的专家以及公共图书馆、高校图书馆、博物馆等图情机构主导，这些专家和图情机构承担着推动数字人文基础设施建设及发展的重任，数字人文自然也成为图

情领域的研究热点，同时也催生了图情学科与其他学科之间的跨学科交叉课题项目出现。

3.4.1 图博档领域

图书馆、博物馆、档案馆作为伴随人类文明进步的公共文化记忆机构，有着保护文化遗产、保存与提供信息资源等天然使命[24]。因此，数字人文项目的不断推进在图书情报领域进行了大量的基础数据积累、综合性工具平台建设以及数字人文教育工具构建。在基础数据积累方面，复旦大学建设的"多源异构学术成果大数据的融合与揭示项目"提出的学术成果大数据的解决方案包括从数据获取到清洗再到关联挖掘及展示的全程处理机制，设计建立可对关键环节实现人工校准机制的交互式数据清理流程。上海图书馆的"SinoPedia:关联数据服务平台"降低了关联数据使用的技术门槛，通过便捷的文件配置方式实现多数据集的语义搜索（SOOOPA）、数据发布（Linked Data Publishing，LDP）和图谱呈现（Linked Data Visualization，LDV）。最后，学科服务是图书馆支撑学科建设的基础服务，因此，在数字人文的背景下，华东师范大学的分布式数字人文研究与教学实训环境以数字方志为研究内容，将资源与研究工具、研究方法、教学训练、方法学习等集于一个逻辑平台上，构建分布式数字人文研究与教学实训环境。

3.4.2 交叉学科

数字人文本身就是一个交叉学科，图情领域研究者是数字人文与其他学科之间的沟通桥梁，数字人文项目的开发研究采取其他学科的项目课题、基础数据和技术手段，搭建了跨学科的数字人文平台。"战疫中的复旦"专题数字特藏平台是复旦大学图书馆正在开发的项目，旨在全面收集、保存、整理和展示全校各领域在战疫期间形成的各类史料，从医学的角度开发和整合各类战疫专题资料并建立数字特藏平台。复旦大学的东亚语言数据中心将语言学与数字人文学科相结合，对海量语料进行数字化，结合大数据与地理信息系统等数字人文方法，建立了国内最大规模、免费公开、面向学术研究的方言语料库。

4 技术手段与方法应用:数字人文项目的研究方法

学科的发展离不开科学技术的支撑，随着数字人文研究对象的多样化和研究内容的跨学科性发展特点，数字人文研究在收集、发现、协作、比较、发布等阶段都离不开文本分析、社会网络分析、空间分析、自动分类、互动测量等技术手段的交叉使用。数字人文研究按技术类型可以分为数字化技术、数据管理技术、数据分析技术、可视化技术、VR/AR技术、机器学习技术等[2]。这些技术实现了数字人文项目的基础资源加工、服务系统构建、应用平台构造、直观形象呈现、交互环境打造以及智慧服务提供。表4是本次51个数字人文项目采用的方法和技术统计情况：

表4 数字人文项目方法和技术

技术类型	具体方法与技术	项目名称
数字化技术	扫描、拍摄、3D建模、高清扫描	数字敦煌、中国地方历史文献数据库（契约文书）、云南省古籍数字图书馆、再现圆明园、南京地区侵华日军慰安所的AR故事地图
数据管理技术	本体构建、数据库设计、命名实体识别、语义搜索、API数据服务、语义描述	《中国历史人物资料库》（CBDB）在线查询系统第二版——面向用户需求的重新设计与实现、融合机器智能与群体智慧的古籍数据化实践、分布式数字人文研究与教学实训环境、敦煌壁画主题词表及关联数据发布服务平台、家谱知识服务平台、SinoPedia:关联数据服务平台、中国知网数字人文研究平台、历史人文大数据平台、地方志数据库计划、民国时期期刊语料库、中国历代人物传记资料库（CBDB）
数据分析技术	文本分析、内容挖掘、地理信息系统、社会网络关系	唐宋文学编年地图、董其昌数字人文展示系统、六朝建康城市历史信息系统、南京地区侵华日军慰安所的AR故事地图、方志物产知识库构建及深度利用研究、中国近现代书画印本数据库、丝绸之路地理信息开放平台、家谱知识服务平台、北京记忆、学术地图发布平台、台湾百年历史地图、文化资源地理信息系统、台湾历史文化地图（THCTS）、唐诗别苑、历史人文大数据平台、中国历代人物传记资料库（CBDB）、"战疫中的复旦"专题数字特藏平台
可视化技术	知识地图、场景模拟、信息美学	台大数字人文中心建构之数字人文分析系统与个人DH研究平台、"战疫中的复旦"专题数字特藏平台、中国古籍基础数据分析平台、宋元学案知识图谱可视化系统、家谱知识服务平台、中国古代皇室家族树（赵宋、李唐、朱明）、宋词缱绻，何处画人间、唐诗别苑、历史人文大数据平台、历史文化名人游学足迹知识图谱的构建与可视化、链上的开放知识图谱联盟
VR/AR技术	虚拟现实技术、增强现实技术、游戏化学习	数字敦煌、再现圆明园、南京地区侵华日军慰安所的AR故事地图、北京记忆
机器学习技术	深度学习、自动分类、个性化服务	方志物产知识库构建及深度利用研究、历史人文大数据平台、"战疫中的复旦"专题数字特藏平台、唐诗别苑、九歌——计算机古诗作诗系统

从表4可以发现当前数字人文项目较多聚焦于数据管理技术、数据分析技术和可视化技术三个类型。换而言之,目前数字人文项目关注的重点已经从单纯的基础资源建设迈向数字人文服务系统、应用平台和直观形象的构建;从初期的数据收集和发现阶段发展成为对数据的收集、发现、协作、比较和发布阶段。这符合数字人文研究的一般生命周期,也意味着信息技术以不同形式深度参与数字人文研究的不同阶段,提供了数字人文研究所需的资源、工具、技术和交流平台。

其次,从表4中可以发现越来越多的数字人文项目已经不单单运用某一类型的技术或是建设某一方面的资源类型,而是逐步发展为多资源、多角度、多方法、多领域的综合性工具与研究平台,旨在为人文专家提供数字人文研究的技术支持和解决方法,在解决传统人文学科问题的同时,也鼓励他们使用平台提供的技术和工具用非人文领域研究方法来理解和分析自有的数据集,并通过工具集成平台分享研究成果[3]。例如,中国中文信息学会语言与知识计算专业委员会主办的"链上的开放知识图谱联盟(Open Knowledge Graph,OpenKG)"是知识图谱数据集的开放共享平台,为用户提供数据库API接口,检索查询系统、预训练模型等工具集,帮助用户构建自己的知识图谱数据。上海图书馆的"历史人文大数据平台"利用语义网、关联数据、知识图谱、机器学习、GIS、可视化等新技术,集合古籍、家谱、名人档案、手稿、近现代图书、报纸、期刊资源等极具特色的馆藏文献,将大规模、全量级、多种类的文献资源汇聚在一起,与人、地、时、事等语义知识库关联,建设数据基础设施、形成多重参照体系,为研究者提供发现新问题的"沃土"、解决老问题的"新工具"。同方知网北京技术有限公司建设的"中国知网数字人文研究平台"也是整合了知网多种文献资源和用户自有资源,通过文献计量分析、文本挖掘分析、可视化呈现、知识图谱构建等研究工具,满足人文研究学者的需求。

5 数字人文与基础设施建设:数字人文项目的类型分析

上海图书馆刘炜研究员等认为数字人文基础设施是一种支持人文科研活动的基础设施,是在数字环境下为开展人文研究而必须具备的基本条件。数字人文基础设施可以分为3个层次,其核心是文献资源和其服务机构组成,提供了基本研究素材的保障;中间层是数字人文研究活动的主体,由系统平台、工具与软件、数据资源建设仓储、基金会、领域专家和数据科学家等构成;外围是数字人文研究的成果发布平台,与学者、用户、社会产生交互,包括数据发现平台、可视化平台、交流平台、开放出版平台等[18]。基于数字人文基础设施"三层次"构成论述,51个项目的层次模块的详细分布归属如表5所示:

表5 数字人文项目基于基础设施建设的类型分析

层次	模块	项目(编号)
核心层	大学、研究所、图档博等机构组织	XM01-XM51的建设机构
中间层	系统平台	XM01、XM04、XM05、XM11、XM13、XM14、XM15、XM16、XM18、XM23、XM26、XM31、XM47
	数据资源机构仓储	XM02、XM03、XM04、XM06-XM12、XM14、XM17、XM19-XM22、XM25-XM30、XM32-XM40、XM42-XM45、XM48、XM49
	计算设施	XM05、XM24、XM41
	领域专家	/
	基金会	/
	管理工具	XM05、XM15、XM23、XM24、XM31、XM32、XM51
外围层	可视化呈现平台	XM03、XM25、XM46、XM50
	数据分析平台	XM46、XM50
	SNS交流平台	/
	开放语义出版平台	/

从表5中可以发现,51个数字人文项目涉及了基础设施建设的方方面面且各有侧重,而这些数字人文项目的工作同样也推动了基础设施的建设,3个层次多个模块因素的相互作用形成了一个可持续发展的有机整体,并且呈现以下3个特征:

(1)数字人文基础设施建设仍然处于"积淀"阶段。本次会议的主题是"积淀与超越",而目前数字人文设施建设也正处于"积淀"阶段。数字人文基础设

施建设是一个长期过程,目前汉语语境下的基础设施建设仍然存在不足,绝大部分的项目主体类型是数字资源建设、系统建设、平台与软件开发。51 个项目中 40 个项目是以数字人文为导向的资源数据库建设,资源数据库的建设即包括将非数字化的人文资料加工转化为数字内容,也包括将非结构化的数字文本内容按照某种使用目的进行规范化标注著录的数据集建设。加工建设的资源类型可分为图像、文本型、音视频以及结构化数据;19 个项目涉及系统建设和平台软件的开发,例如北京大学主持的“《中国历史人物资料库》(CBDB)在线查询系统第二版——面向用户需求的重新设计与实现”项目、复旦大学主持的“多源异构学术成果大数据的融合与揭示”项目等。而只有同方知网主持的“中国知网数字人文研究平台”、上海图书馆主持的“历史人文大数据平台”2 个项目涉及了基础设施建设的三个层次,有望形成可自我运行、不断优化、可持续发展的有机整体,这也代表着汉语语境下数字人文工作在不断“积淀”的基础上进行超越的方向。

(2)数据资源及机构仓储建设是基础设施建设中的重点和热点。在这 40 个涉及数字资源数据库建设的项目中有 22 个项目的数字资源内容覆盖面较窄,属于特藏资源的数字化建设,即专题数据库的建设。专题库的建设对象往往是某一特定领域或特定主题的数据资源,往往是以某一部馆藏典籍为基础,或是某一特定历史时期、历史事件为主题进行相关资源的数字化与数据化建设与服务。例如,南京农业大学的“方志物产知识库构建及深度利用研究”项目是对约 3000 余万字的《方志物产》进行了数字化和语义化加工,构建涵盖词汇、术语、关键词、命名实体、时空序列的知识库。北京大学数字人文研究中心开发的“宋元学案知识图谱可视化系统”则是从遥读、精读两个角度对《宋元学案》这部理学巨著进行深度挖掘和分析,讲述宋代师承和学术思想演变过程。南京大学的“六朝建康城市历史信息系统”以六朝时期的建康(今南京)为时空研究范围,建立了完整的六朝建康历史地理数据库。

(3)数字人文基础设施建设是“使能者”[25]。虽然未收集到 51 个项目的基金赞助情况,但是刘炜等[18]指出各类基金会是数字人文基础设施建设的重要推动力量,同时其自身也是基础设施建设的一部分。而国内也有越来越多的数字人文项目得到了“国家社科基金”、“教育部哲学社会科学研究基金”等资助,从而也吸引到更多的领域专家参与到数字人文项目的建设中来。此外,大学、研究所和图档博等机构作为基础设施建设的核心因素,纷纷成立数字人文研究中心,并以此为建设主体主持开发数字人文项目,例如人民大学数字人文研究中心主持的“高迁古村数字记忆项目”、武汉大学数字人文研究中心主持的“基于 IIIF 的敦煌壁画数字叙事系统”北京大学数字人文研究中心主持的“宋元学案知识图谱可视化系统”,上海图书馆的历史人文大数据平台等。这些机构在做好保障基本研究资料的基础上为各类人文研究项目提供了数据共享、机构协作、人员培训的协助[4],也为学术交流、开放获取、共建共享、学科融合发展提供渠道和平台,推动人文社会科学的定量化和科学化研究。

6 服务模式与问题解决:数字人文项目的现实意义

在 2020 年数字人文年会的专家辩论环节,专家们一致认为数字人文基础设施建设不能取代人文科学本身的研究,数字人文基础设施建设存在的意义是推动人文研究的发展[25]。从上面分析可以看出,目前汉语语境的数字人文项目工作基本都还是属于基础设施建设范畴,也就是说数字人文项目通过为数字人文学者乃至传统人文学者解决科研中的实际问题,提供资源素材保障及工具平台计算设施服务,从而推动人文研究的发展,实现其现实意义,主要体现在以下 3 个方面:

(1)数字人文项目解决学者基本研究素材保障、数据处理和软件学习的实际问题。目前,很多人文学者在做研究时在数据收集处理和软件技术的学习上花费大量时间,项目中大量涉及数字资源建设和工具软件开发,这些都可以为学者提供基本的研究素材,帮助他们进行数据加工处理,将学者从繁复的资料整理及数据处理的工作中解脱出来,更好投入到本学科领域的研究中,提高研究效率,促进学科发展。此外,吸引人文学者一起参与数字人文项目工作就可以在项目的各个阶段充分了解人文研究工作的需求,建设更符合人文研究学者需要的资源数据库、工具软件和系统平台。

(2)数字人文项目提供学者开放数据应用、个性化研究支持的服务。一方面,一些项目开发的工具或平台、建设的开放数字资源可以被其他项目利用和进一步开发、优化,完成数字资源建设的不断迭代,有效促进信息资源的二次利用,创造出新的“数据价值”,为人文学者提供基于“数据驱动”的服务。例如上海

大学的"历史文化名人游学足迹知识图谱的构建与可视化"项目、清华大学的"中国古代皇室家族树(赵宋、李唐、朱明)"等项目都是利用已有开放数据的成果进行艺术和历史领域的再研究,两个项目都在研究中引入 CBDB 的人物数据,让传统的历史人文研究焕发新光彩,也为艺术、历史领域人文学者提供新理念和新角度。另一方面,一些平台和系统支持人文学者的个性化研究,进行处理数据、建数据库,结合不同来源的工具与数据集等工作。例如,台湾大学的"台大数字人文中心建构之数字人文分析系统与个人 DH 研究平台"就提供工具处理个人资料、一键建立数据库、分析与可视化工具及个人语料库服务。"复旦大学东亚语言数据中心"项目开发的自定义语言条件语言地理信息查询、用户上传数据制图系统等都满足了专家学者的个性化使用需求。这些平台和系统极大方便了研究者的科研活动,为人文研究学者提供一系列人文学科领域相关的资源、工具和服务,帮助学者建立个人学术档案,使平台成为分享和传播数字人文信息的桥梁,满足人文学者研究的多样性需求。

(3)数字人文项目具有助力新文科建设、促进中华文化传承的现实意义。一方面,随着《新文科建设宣言》的发布,我国人文社科研究正式进入全新的历史时期,"第四范式"的数据驱动型研究在人文研究领域获得广泛推崇。王丽华、刘炜[26]以"数字人文堆栈 2.0"的理论模型为新文科提供了颠覆性的学科范式,指出数字人文从机构、组成、方法、系统和界面五个层次以及实体与精神的两个维度为新文科建设提供了相应的助力点和借鉴元素。而数字人文项目的成果,例如数据集、工具、系统、可视化作品等,都可以视为在新文科建设背景下数字人文乃至新文科成果交流的一种重要呈现方式。另一方面,数字人文项目建设也促进了中华文化的传承。正如本届会议的主题"积淀与超越:数字人文与中华文化",汉语语境下的数字人文项目应该包含以"仁"为中心的中国人文主义滥觞,以及以儒家"六艺"为代表的人文学科和精神积淀,从而具有其独特的内涵。不少项目在历史文化研究方面有成功的实践经验,例如中国人民大学主持的"高迁古村数字记忆网站 "项目、武汉大学建设的"基于 IIIF 的敦煌壁画数字叙事系统"项目和"敦煌壁画主题词表及关联数据发布服务平台"项目、北京大学参与的"中国历代人物传记资料库(CBDB)"项目以及中南民族大学开发的"唐宋文学编年地图"项目等。这些项目,用数字人文的理论、方法和技术保护、传承、构建及呈现中华文化记忆,使中华传统文化得到传承,为中华文化研究做出了贡献。

7 结语

通过对 51 个项目进行梳理和综述,发现汉语语境下的数字人文项目主要发展趋势如下:

(1)史学和文学领域是数字人文研究蓬勃发展的学科领域。在这两个学科领域中,地理信息系统、社会网络分析、文本分析等技术被广泛应用。语言艺术学科是数字人文项目建设的新领域,数据可视化技术以及优秀的视觉传达也受到数字人文项目建设者的诸多关注。

(2)机器学习、VR/AR 技术与数字人文结合空间巨大。作为新型的信息技术方法,VR/AR 技术和机器学习(特别是人工智能技术)开始被引入数字人文项目中,衍生出了自动作诗、古籍自动识别、故事地图、游戏化学习等新的研究方向和应用场景,为人文学者和普通用户都提供了新的服务和体验,虚拟现实和机器学习会是数字人文研究的新热点。

(3)数字资源建设在数字人文基础设施建设中仍然是重中之重,这也是图档博建设者的天然使命。而其中特藏资源的专题数据库建设又是数字人文基础设施建设的重点工作。数字资源的建设对象和建设内容也从单一的文本数据,发展为非结构化数据、音视频、图片数据的建设。

(4)综合性数字人文平台开发与建设的时代已经到来。项目开发方在某一类型资源或工具开发的基础上集结历年开发经验和基础设施资源,开始进行集成多种数字资源统一服务、多工具共享、提供人文学者研究合作、成果发布服务等为一体的综合性数字人文平台。上海图书馆的"历史人文大数据平台"、中国知网的"数字人文研究平台"都是其中的领航者。

(5)跨学科、跨机构、跨国家的合作性项目开始显现。数字人文作为一个新兴的交叉学科与文学、历史学、艺术、语言学、计算机科学等传统学科呈现出研究方法交叉融合、内容理念交流深入、研究成果多样展现的新趋势。同时,项目的建设方也呈现了跨机构、跨地区、跨国家的项目交流合作模式。

同时,研究也发现当前数字人文项目的建设和发展存在以下不足,未来数字人文项目的开展需要进行针对性的布局谋篇:

(1)文献资源数据化程度不足。包括了文本的获取不充分、版权垄断和开放获取程度不足等,非结构化

的文本资源限制了人文学者的研究也制约了数字人文项目的开发深度。继续加强数字资源基础设施建设的同时，应鼓励将数字化资源开放给学者，加强机构间合作，避免重复建设。

（2）多学科交叉融合度仍然不足。虽然跨学科的趋势开始呈现，但是数字人文学科的边界仍然模糊，从而导致了人文研究数字化和数字技术人文化两个维度未能很好的融合，未能在人文研究与数字技术的跨学科结合中起到桥梁作用。因此，应当加强多学科的交叉融合，使更多的领域研究专家（领域专家、数据科学家、算法工程师）参与到数字人文项目的建设中来。

（3）研究工具和数据运用的科学训练不足。虽然已经出现综合性数字人文平台的建设，但是仍然缺少综合性、系统性的数字人文培训的平台和案例。各机构需要加强大数据交叉思维和人文素养的培养，引发学者对数字新技术和新展现方式的探索与思考。

（4）数字人文基础设施建设和数字人文研究发展不匹配。继续提升学科服务能力和基础资源建设，基础设施建设要与时俱进并且依靠各个学科的创新。同时拓展新的人文领域研究范畴，响应教育部新文科建设的号召，促进数字人文研究领域理论的体系化、成熟化。

（5）区域发展不平衡。一线城市机构主持建设的数字人文项目占比超过半数，上海、北京是国内数字人文研究的领先地区。中东部地区数字人文项目建设水平高于西部地区，江苏、浙江、湖北是国内数字人文项目建设的第二梯队区域。港澳台地区特别是台湾地区十分重视数字人文领域的建设与投入。区域发展的不平衡既是经济和社会发展存在差异的体现，更是数字人文项目建设主体即高校和图档博机构的建设意识和实际投入不均衡所导致。会议期间发起成立的中国索引学会数字人文专业委员会，工作成效显著，成员单位发展势头迅猛，短短半年已经有云南大学历史与档案学院数字人文工作室等13家数字人文机构加入。发展和团结更多不同地区的数字人文机构和组织加入数字人文专业委员会，推动国内其他如中西部地区的数字人文项目及相关领域工作的研究探索，加强区域间协同发展也是专委会的工作目标之一。

参考文献：

[1] KIRSCHENBAUM M G, What is digital humanities and whatś it doing in English Departments? [EB/OL]. [2021-08-30]. https://www.uvic.ca/humanities/english/assets/docs/kirschenbaum.pdf.
[2] 刘炜，叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报，2017，43(5)：32-41.
[3] 朱本军，聂华. 跨界与融合：全球视野下的数字人文——首届北京大学"数字人文论坛"会议综述[J]. 大学图书馆学报，2016，34(5)：16-21.
[4] 高胜寒，赵宇翔，朱庆华. 国内外数字人文领域研究进展分析[J]. 图书馆杂志，2016，35(10)：9-18.
[5] Yale university library digital humanities lab [EB/OL]. [2021-07-13]. http://dhlab.yale.edu.
[6] Hyperstudio-digital humanities at MIT [EB/OL]. [2021-07-13]. http://hyperstudio.mit.edu.
[7] 徐彤阳，顾婷婷. 日本数字人文项目探析和启示[J]. 国家图书馆学刊，2021，30(3)：88-99.
[8] Sydney digital humanities research group [EB/OL]. [2021-07-13]. https://www.sydney.edu.au/arts/our-research/centres-institutes-and-groups/sydney-digital-humanities-research-group.html.
[9] 邓要然. 美国图书馆领域数字人文项目研究[J]. 图书馆工作与研究，2017(8)：29-35.
[10] 赵雪芹，莫长镭，雷春蓉. 美国高校图书馆数字人文项目调研与启示——以美国排名前10位的高校为例[J]. 图书馆，2021(1)：70-76.
[11] 徐彤阳，杨明睿. 澳大利亚数字人文项目透视——兼论对我国数字人文项目的启示[J]. 图书情报工作，2020，64(22)：145-152.
[12] 林泽斐. 英国数字人文项目研究热点分析——基于DHCommons项目数据库的实证研究[J]. 情报资料工作，2018，39(1)：97-104.
[13] 蔡迎春. 特色资源建设中的数字人文应用进展研究——基于国内数字人文相关项目及实践案例[J]. 图书馆建设，2018(7)：18-24，30.
[14] 苏敏. 图书馆开展数字人文服务研究综述[J]. 情报探索，2020，4(1)：120-126.
[15] 吴丽萍. 我国图书馆数字人文研究现状及展望[J]. 图书馆工作与研究，2021，4(6)：30-36.
[16] 朱本军，聂华. 互动与共生：数字人文与史学研究——第二届"北京大学数字人文论坛"综述[J]. 大学图书馆学报，2017，35(4)：18-22.
[17] 吕璐成，韩涛. AI在图情：人工智能赋能图情服务——2019年图书馆前沿技术论坛(IT4L)会议综述[J]. 农业图书情报学报，2020，32(5)：13-18.
[18] 刘炜，谢蓉，张磊，等. 面向人文研究的国家数据基础设施建设[J]. 中国图书馆学报，2016，42(5)：29-39.
[19] 夏翠娟. 面向人文研究的"数据基础设施"建设——试论图书馆学对数字人文的方法论贡献[J]. 中国图书馆学报，2020，46(3)：24-37.
[20] 中央文化体制改革和发展工作领导小组. 关于做好国家文化大数据体系建设的通知[EB/OL]. [2021-02-01]. https://www.gujianchina.cn/news/show-9194.html
[21] 赵宇翔，练靖雯. 数字人文视域下文化遗产众包研究综述[J].

数据分析与知识发现,2021,5(1):36－55.

[22] 曾熙,谭旭,王晓光. 文化遗产大数据二维分类框架研究[J]. 图书情报知识,2020,37(1):84－93.

[23] 吴健. 多元异构的数字文化——敦煌石窟数字文化呈现与展示[J]. 敦煌研究,2016(1):123－127.

[24] YARROW A, CLUBB B, DRAPER J. Public libraries, archives and museums: trends in collaboration and cooperation[R]. The Hague: IFLA Headquarters, 2008.

[25] 颜佳,姚啸华. 数字人文发展的"主导者"与"使能者"——2020 数字人文年会"数字人文基础设施建设"专家论辩综述[J]. 数字人文,2021(1):123－133.

[26] 王丽华,刘炜. 助力与借力:数字人文与新文科建设[J]. 南京社会科学,2021, 4(7):130－138.

作者贡献说明:

沈立力:负责收集资料,撰写论文;

张宏玲:负责论文选题、提出研究思路、确定框架、修改论文;

韩春磊:负责修改论文;

刘炜:负责修改论文。

Analysis and Enlightenment of Digital Humanities Project in Chinese Context

——The Overview of the Project Selection of DH2020

Shen Lili　Zhang Hongling　Han Chunlei　Liu Wei

Shanghai Library, Shanghai Institute of Science and Technology Information, Shanghai 200031

Abstract: [**Purpose/significance**] This paper analyzes 51 digital humanities projects in Chinese context which is collected at the 2020 digital humanities conference held by Shanghai Library. This paper summarizes the general situation of digital humanities projects on Chinese cultural themes in the Chinese context, and provides reference for the future development of digital humanities in China. [**Method/process**] The 51 projects were analyzed by network survey and data analysis, and classified according to the different research objects and contents. Meanwhile, valuable features were obtained by analyzing and discussing the research methods, implementation conditions and practical significance of the project results. This paper discussed four dimensions of "Digital Humanities and humanities" "Application of technical means and methods" "infrastructure construction" and "extended service and problem solving". [**Result/conclusion**] The overview of 51 projects show five characteristics: Digital Humanities is flourished in the fields of history and literature, and has deep combination of machine learning. The construction of digital resources is still the top priority. The era of Comprehensive Digital Humanities platform has come, and interdisciplinary, inter institutional and international cooperative projects have also begun to appear. In view of the shortcomings in the construction and development of Chinese Digital Humanities projects, this paper also puts forward the main suggestions to promote inter agency cooperation and interdisciplinary integration, to strengthen the cultivation of big data cross thinking and humanistic quality, and continue to improve the subject service ability and basic resources construction.

Keywords: digital humanities　DH2020　digital humanities annual conference　Chinese culture　project analysis